

Supplementary Information

A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing

Beverly Y. Mok, Marcos H. de Moraes, Jun Zeng, Dustin E. Bosch, Anna V. Kotrys, Aditya Raguram, FoSheng Hsu, Matthew C. Radey, S. Brook Peterson, Vamsi K. Mootha, Joseph D. Mougous and David R. Liu

	Page #
Supplementary Discussion	3-7
Supplementary Data 1. Sequencing coverage of ATAC-seq samples	8
Supplementary Data 2. Number of overlapping off-target SNVs	9
Supplementary Data 3. Dual fluorescence micrographs and uncropped gel images for Extended Data Fig. 6f	10
Supplementary Data 4. Uncropped images for Extended Data Fig. 6h	11
Supplementary Table 1. Diffraction data collection and refinement statistics for DddA / DddI _A	12
Supplementary Table 2. Schematic and sequences of guide RNAs for split-DddA _{tox} -Cas9 screen	Excel
Supplementary Table 3. Base percentages at each position of the <i>EMX1</i> locus for DddA _{tox} -Cas9 splits with no guide RNAs and monomers with their respective gRNAs	Excel
Supplementary Table 4. List of mitoTALE binding sites for indicated DdCBE	15
Supplementary Table 5. Indel frequencies of DdCBEs in their optimized orientations	Excel
Supplementary Table 6. Base percentages at each position of the H-strand of mtDNA for indicated DdCBE in its optimized split orientation	Excel
Supplementary Table 7. P-values from comparison of editing efficiencies from time course experiments in HEK293T cells and U2OS cells	Excel
Supplementary Table 8. Unique off-target SNVs mediated by DdCBEs	Excel

Supplementary Table 9. Overlapping off-target SNVs	Excel
Supplementary Table 10. Disease-associated mitochondrial DNA point mutations	21-22
Supplementary Table 11. Bacterial strains and plasmids used in this study	23-24
Supplementary Table 12. Sequences used for DddA _{tox} characterization in bacteria	25-27
Supplementary Table 13. Sequences used for characterization of DddA _{tox} and its fusions in mammalian cells	28-31
Supplementary Sequences 1. Intact DddAtox–Cas9 fusions sequences	32-33
Supplementary Sequences 2. Split-DddAtox–Cas9 fusions sequences	34-35
Supplementary Sequences 3. Split-DddAtox–CCR5 TALE fusions sequences	36
Supplementary Sequences 4. General DdCBE architecture and mitoTALE sequences	37-41

Supplementary Discussion

Toxicity of intact DddA_{tox} fused to DNA-binding proteins

To investigate DddA_{tox} activity in human cells, we transfected HEK293T cells with plasmids encoding DddA_{tox} fused to catalytically inactive *S.pyogenes* Cas9 (dSpCas9)^{23,58} or SpCas9(D10A) nickase^{58,59} (**Supplementary Sequences 1**). We observed a 2-10 fold lower viability in cells transfected with DddA_{tox}-Cas9 fusions compared to BE2max (a non-nicking cytosine base editor with two copies of UGI protein)¹⁴ and BE4max (an optimized current-generation cytosine base editor)^{28,30} (**data not shown, available upon request**). The widespread toxicity induced by DddA_{tox}-Cas9 precluded PCR amplification of target DNA sites from the few surviving cells for high-throughput DNA sequencing.

Editing efficiencies of DddA_{tox} splits G1322, A1343, N1357, G1371 and N1387

Splits A1343, G1371, and N1387 resulted in <0.1% conversion of C•G-to-T•A at TC target bases across all tested spacing lengths (**Extended Data Fig. 3**). G1322 and N1357 gave a maximum of 8.3% and 22% C•G-to-T•A conversion, respectively, within a 44-bp spacing regions (**Extended Data Fig. 3g**), but base-editing efficiencies at other lengths of spacing regions were generally below 5.6% (**Extended Data Fig. 3**).

Indel frequencies associated with split-DddA_{tox}-Cas9 fusions

The level of indels associated with splits that afforded the highest editing efficiency (G1333 and G1397) were higher than those of canonical cytosine base editors³⁰, up to 46% for some split fusion combinations (**Extended Data Fig. 3**). We speculate that strand nicking by SaKKH-Cas9(D10A), combined with uracil excision on either strand¹⁹ can induce double-strand breaks that induce indel formation through end-joining processes^{60,61}.

Preliminary observations governing editing efficiency and selectivity of split-DddA_{tox} for subsequent DdCBE editing

Cytidines in TCC contexts appeared to be favorable substrates for deamination by split-DddA_{tox}. TCC target bases are generally deaminated at varying efficiencies depending on their positions within the dsDNA spacing region (**Extended Data Fig. 3d-h**). We observed modest C•G-to-T•A conversions between 2.0-8.5% within spacing region lengths of 28-, 33- and 39-bp (**Extended Data Fig. 3d-f**). Editing efficiencies were elevated to 48% and 26% at 44- and 60-bp spacing lengths, respectively (**Extended Data Fig. 3g, h**). In addition to TCC,

split-DddA_{tox} also mediated deamination of cytidines in TCA (**Extended Data Fig. 3a-c, e-g**) and TCT (**Extended Data Fig 3b,c, g**) contexts.

The selectivity of editing also depended on the fusion orientation. For 17-bp spacing length, we observed preferential deamination of target TC's that were closer to the protospacer of the Cas9 monomer fused to DddA_{tox}-C. (**Extended Fig. 3b**). For longer spacing region lengths (44- and 60-bp), target bases closer to the protospacer of the Cas9 monomer fused to DddA_{tox}-N were preferentially deaminated (**Extended Data Fig. 3g, h**).

In light of the above results, we designed subsequent mitoTALE binding sites to flank a spacing region that contains TCC and/or TCA target bases. To assess the substrate selectivity and editing efficiency of mitoTALE–split-DddA_{tox}, we also tested the two possible orientations for each DddA_{tox} split at G1333 and G1397; DddA_{tox}-N is fused to the right-side TALE and DddA_{tox}-C is fused to the left-side TALE, or DddA_{tox}-N is fused to the left-side TALE and DddA_{tox}-C is fused to the right-side TALE.

Cas9-DNA interactions are structurally distinct from TALE-DNA interactions. As such, spacing region lengths that are amenable to deamination by DddA_{tox}–Cas9 may not apply to that of TALE–DddA_{tox} fusions. However, the programmability of Cas9 proteins enabled rapid testing of many different spacing lengths in cells to identify split sites that produce DddA_{tox} halves which can reassemble into a catalytically active deaminase. Interrogating different spacing region lengths *in vivo* with TALE proteins would be much more labor-intensive, and could be confounded by differences in TALE binding affinities for their target sites

TALE and ZFP fusions of APOBEC/AID cytidine deaminases

While cytidine deaminases have previously been fused to zinc-finger arrays and TALE arrays to recognize a single target sequence within a mammalian genome, the very low activity of previously described deaminases on dsDNA resulted in <2.5% editing efficiencies in human cells that are spread over >150 bp around the DNA-binding site³⁵.

Mitochondrial DNA repair and editing outcomes

The very high product purity and lack of indels from mtDNA editing by DdCBE supports a model in which lesion-containing genomes are degraded rather than repaired⁶². This model is consistent with previous observations that a ssDNA deaminase targeted to the mitochondria introduced random C•G-to-T•A conversions but no indels⁶³. The result is selective maintenance of mtDNA copies that have been cleanly edited.

Among the mixture of edited alleles produced by ND6-DdCBE, AC₁₁-to-AT₁₁ conversion was present only in alleles that already had edits at the more readily edited 5'-TC contexts on the bottom strand (equivalent to 5'-GA-3' on top strand) (**Fig. 3e**), suggesting that DddA_{tox} may operate processively. Since this behavior was not apparent in our bacterial genome mutational analyses (**Fig. 1i**), it may arise from tethering DddA_{tox} to a DNA-binding protein.

Durability of C•G-to-T•A mitochondrial DNA conversions in HEK293T cells

To evaluate the durability of C•G-to-T•A mitochondrial DNA conversions in HEK293T cells, we tracked editing induced by *ND6*-, *MTND5P1*-(denoted as *ND5.1*-), *MTND5P2*-(denoted as *ND5.2*- and *ATP8*-DdCBE over 18 days at 3- or 6-day intervals, spanning approximately 21 cell divisions. Across all DdCBEs, editing efficiencies increased by 1.5- to 3.7-fold from day 3 to day 6 as levels of DdCBE protein persisted (**Extended Data Fig. 6a-e**). DdCBE proteins were not detected by day 12 (**Extended Data Fig. 6f**), but editing levels were generally maintained and sustained through the end of the experiment (day 18). By comparison, nuclear base editing of *EMX1* by BE4max and its non-nicking equivalent (BE2max)^{14,30} peaked at day 3 and was generally maintained through day 18 (**Extended Data Fig. 6e**). From day 12 to day 18, cells treated with *ND6*-DdCBE had a 6.2% decrease out of 27-34% total edits at C₆ and C₇ (**Extended Data Fig. 6a**). Missense mutations induced by DdCBEs could have impaired mitochondrial fitness, thereby placing them at a selective disadvantage for mtDNA replication. We also noted that BE4max-mediated editing of C₅ and C₆, which decreased by 12% out of 57-70% total editing (**Extended Data Fig. 6e**).

Detecting mtDNA deletions and mtDNA copy number in DdCBE-treated human HEK293T cells

Long-range PCR of the whole mitochondrial genome resulted in a shorter DNA band for *ND6*-DdCBE-treated cells compared to amplicons obtained from cells treated with dead *ND6*-DdCBE (**Extended Data Fig. 6h**). The truncated amplicon from positions 2478-10858 of the mitochondrial DNA restored to its full length after day 12 (left panel), while the truncated amplicon from positions 2688-10653 (right panel) persisted to the end of the experiment (day 18). However, ATAC-seq results revealed no large deletions in HEK293T cells 3 days after treatment with *ND6*-DdCBE, which would have manifested as regions in the mtDNA with no sequencing coverage (**Supplementary Data 1**). We speculate that the shorter amplicons observed in the DNA gel may be a result of PCR artefact. We did not observe a decrease in

mtDNA copy number in cells treated with the indicated DdCBE 3 days post-transfection. It is possible that mtDNA replication could have obscured reductions in copy number (**Extended Data Fig. 6i**).

Design of TALE repeat arrays

De novo TALE-binding sequences were designed using the following guidelines⁶⁴: (i) 14-18 bp long, (ii) thymidine at the 3'-end, (iii) large %C within the first 5 nucleotides, if possible and (iv) large %T within the last 5 nucleotides, if possible. The RVDs used in this manuscript to recognize A, C, T and G/A nucleotides are NI, HD, NG and NN, respectively^{65,66}. It should be noted that the NH and NK repeats have been reported to recognize only G, but TALE activity of NK repeat is lower than that of NN repeat⁶⁷. The C-terminal TALE domain of each mitochondrial-targeting DdCBE reported in this manuscript is 41-amino acids long. DddA_{tox} half is fused to the TALE protein by a 2-amino acid linker. The average mtDNA editing efficiencies were highest for 14- and 16-bp spacings (**Fig. 4a-c**). These spacing length preferences are consistent with those of TALENs containing a 39-amino acid long C-terminal domain fused to *Fok I* by an 8-amino acid linker, and with 12-24 bp spacings supporting the higher nuclease activity⁶⁸.

We noted that the following mitoTALEs (*ND6-right*, *ND6-left*, *ND5.2-right* and *ATP8-left*) contained a mismatched terminal TALE repeat^{9,69}. Imperfect binding of mitoTALEs to off-target mtDNA site could account for the lower editing efficiencies observed at these sites compared to those observed from *de novo*-generated DdCBEs, except for *ND2-DdCBE* (Compare **Fig. 4e, f** to **Fig. 4a, c**).

Characterizing mitochondrial function of edited HEK293T 6 days after transfection

We set out to evaluate the effects of DdCBE editing at *MT-ND5* and *MT-ND1* on mitochondrial function (**Extended Data Fig. 8c-f**). We confirmed that editing by *ND5.1-* and *ND5.3-DdCBE* did not significantly alter oxygen consumption rates and other respiratory parameters in HEK293T cells, as expected from the silent C-to-T mutations (**Extended Data Fig. 8c, e**). Editing by *ND5.2-* and *ND1-DdCBE* resulted in a point mutation from Ser to Phe and Leu to Phe, respectively (**Extended Data Fig. 8d, f**). In both instances, we observed negligible effects on mitochondrial function. Collectively, these results suggest that DdCBEs could be used to generate targeted point mutations in mtDNA that enable edited cells to be interrogated for phenotypic effects. A mixture of low- and high-efficiency-DdCBEs will be

useful for creating a range of heteroplasmy levels to investigate the minimum threshold for phenotype change and clinical penetrance of known and newly reported mtDNA mutations.

Profile of off-target SNVs

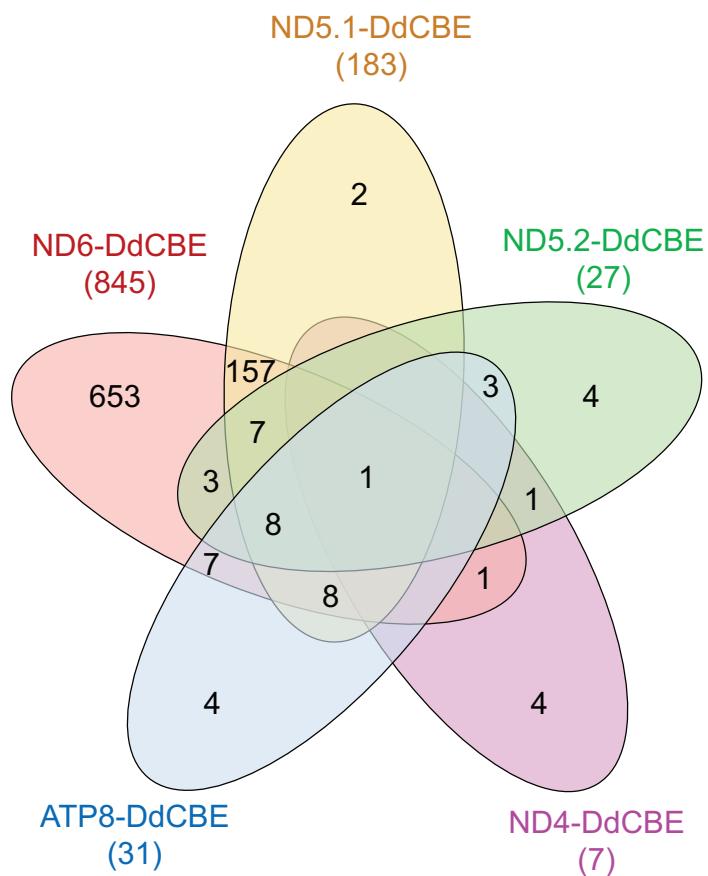
More than 99% of the off-target SNVs in DdCBE-treated samples were C•G-to-T•A transitions while SNVs in dead-DdCBEs and untreated controls were a mixture of DNA transitions and transversions (**Supplementary Table 8**), suggesting that these off-target SNVs arise from DdCBE-mediated deamination rather than cell-specific heteroplasmy or somatic mutations.

Predicted effects of off-target SNVs on mtDNA sequence and protein function

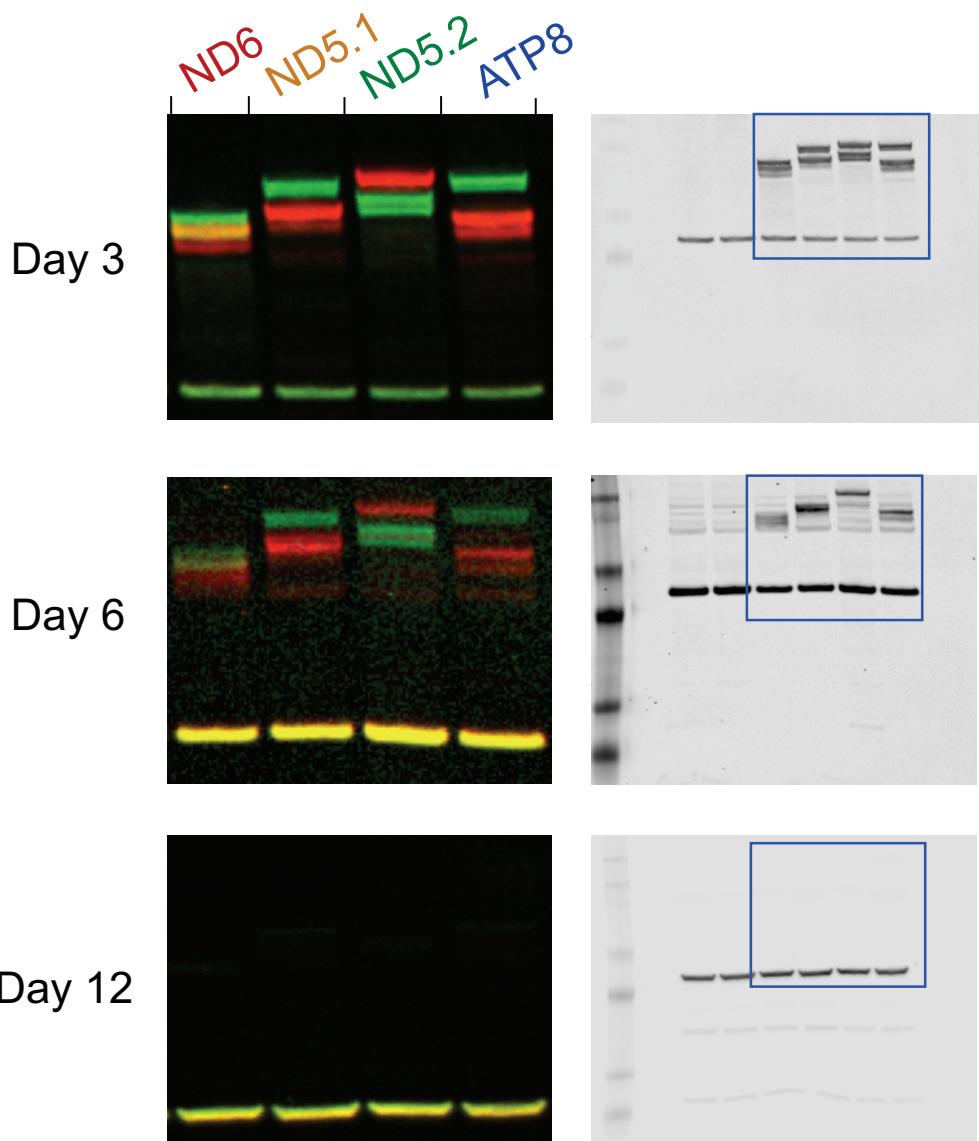
Approximately one-third of the total off-target mutations for each DdCBE resulted in missense mutations, while approximately two-thirds of the off-target mutations were in non-coding regions or led to synonymous mutations in coding regions (**Extended Data Fig. 11 Extended Data a**). Among the missense mutations, more than half were predicted to be deleterious to protein function (**Extended Data Fig. 11b**). We note that unannotated SNVs in non-coding rRNA and tRNA, which were excluded in SIFT analysis, could also have functional consequences⁷⁰. To validate the SNV-calling pipeline, we selected up to 18 SNVs called for the different DdCBEs for targeted amplicon sequencing and verified a total of 35 SNVs to be bona-fide off-target SNVs (**Extended Data Fig. 11c**).



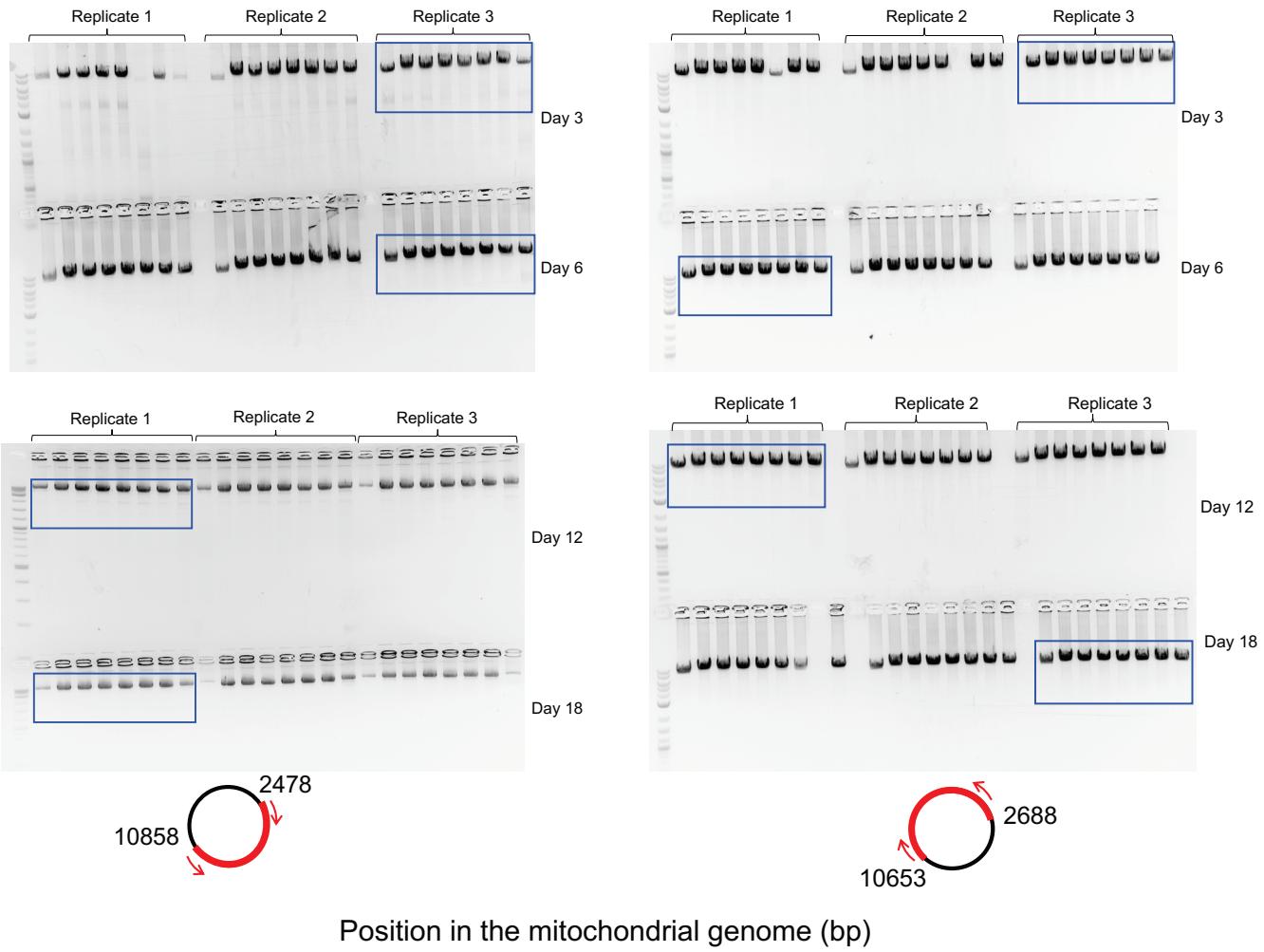
Supplementary Data 1 | Sequencing coverage of ATAC-seq samples. Per-base sequencing coverages of each replicate treated with DdCBEs, dead-DdCBE, and untreated control. The nucleotide positions of the human mitochondrial DNA from the NC_012920 reference genome are indicated in the exterior of each radial plot. Inner red circle represent 5,000x coverage.



Supplementary Data 2 | Number of overlapping off-target SNVs. Venn diagram depicting the number of off-target SNVs shared among two or more DdCBEs. The number of unique SNVs is indicated in parenthesis.



Supplementary Data 3 | a, Dual fluorescence imaging of *SOD2* MTS–left TALE–split DddA_{tox}–UGI half (green) and *COX8A* MTS–right TALE–split DddA_{tox}–UGI half (red) for each DdCBE (see **Extended Data Fig. 6f**). Day 3 images are representative of $n=3$ independent biological replicates. $n=1$ for Day 6 and Day 12 images. **b,** Uncropped images for **Extended Data Fig. 6f**.



Supplementary Data 4 | Uncropped images for Extended Data Fig. 6h.

Supplementary Table 1 | Diffraction data collection and refinement statistics for DddA / Dddl_A.

DddA / Dddl _A SAD peak ^a	
PDB accession code	6U08
Data Collection	
Space group	P2 ₁ 2 ₁ 2
Cell dimension <i>a, b, c</i> (Å) α, β, γ (°)	126.8, 145.0, 64.2 90, 90, 90
Wavelength (Å)	0.9790
Resolution (Å)	38.6-2.5 (2.53-2.50) ^b
No. unique reflections	41658 (1339)
R _{merge}	0.12 (1.2)
I/σI	26.7 (2.0)
Completeness (%)	
Total	98.9 (96.7)
Anomalous	99.0
Redundancy	13.0 (12.6)
Wilson B-factor (Å ²)	46.2
Refinement	
Resolution (Å)	38.6-2.5 (2.53-2.50)
No. reflections	41655 (2789)
Rwork/Rfree (%)	17.2/22.8 (26.3/34.6)
No. atoms	7889
Protein	7796
Ligand/ion	4
Water	89
B-factors (Å ²)	
Protein	59.2
Ligand/ion	72.3
Water	50.7
rmsd	
Bond lengths (Å)	0.012
Bond angles (°)	1.091
Missing residues	
chain A	1250-1289, 1423-1427
chain B	71-3
chain C	1250-1289, 1423-1427
chain D	72-3
chain E	1250-1290, 1423-1427
chain F	none
chain G	1250-1290, 1423-1427
chain H	none

^aAll data collected from a single crystal

^bValues in parentheses are for the highest resolution shell

(See the provided Excel file).

Supplementary Table 2 | Schematic and sequences of guide RNAs for split-DddA_{tox}–Cas9 screen. dSpCas9 guide RNAs (spG7 and spG6) are paired with SaKKH guide RNAs (saG1 to saG4) to generate spacing regions with lengths between 12 and 60 bp.

(See the provided Excel file).

Supplementary Table 3 | Base percentages at each position of the *EMX1* locus for DddA_{tox}-Cas9 splits with no guide RNAs and monomers with their respective gRNAs. For G1333/G1397 DddA_{tox}-dSpCas9 or G1333/G1397 DddA_{tox}-SaKKH-Cas9(D10A) halves, these halves are directed to a site within *EMX1* by a guide RNA spG4 (magenta) or saG4 (blue), respectively. The reciprocal DddA_{tox} half of each fusion was absent. The 60-bp spacing region is highlighted in yellow. All shorter spacing regions are nested within the 60-bp spacing. Base percentages are obtained 3 days post-transfection and representative of *n*=2 independent biological replicates.

Supplementary Table 4 | MitoTALE binding sites for each DdCBE. Sequences start from the 5' nucleotide recognized by the N-terminal domain in the TALE array (N_0 position) and end with the nucleotide recognized by the carboxy-terminal truncated 'half' repeat. See the **Supplementary Discussion** for the *de novo* design of mitoTALEs (blue). Previously published mitoTALE sequences are shown in green. ND6 right-mitoTALE contained a mutant N-terminal domain⁷¹ that enabled recognition of cytosine at the N_0 position (red). See **Supplementary Sequences 4** for mitoTALE protein sequences.

DdCBE	Left-mitoTALE target sequence	Right-mitoTALE target sequence
ND1	5'-T CTAGCCTAGCCGTTT-3' <i>De novo</i>	5'-T GAGTTGATGCTACCCCT-3' <i>De novo</i>
ND2	5'-T CTTAGCATACTCCTCAAT-3' <i>De novo</i>	5'-T AGAACTGCTATTATT-3' <i>De novo</i>
ND4	5'-T GCTAGTAACCACGTTCT-3' <i>De novo</i>	5'-T CCTGTAAGTAGGAGAGT-3' <i>De novo</i>
ND5.1	5'-T AGCATTAGCAGGAAT-3' From Hashimoto. M <i>et al.</i> , 2015	5'-T CTTGGAGTAGAA-3' Adapted from Hashimoto. M <i>et al.</i> , 2015 to recognize C over T at the N_1 position
ND5.2	5'-T CAAAACCATAACCTCT-3' <i>De novo</i>	5'-T GCTAGGCTGCCAATGT-3' ^a From Bacman. S <i>et al.</i> , 2013 (previously denoted as $\Delta 5$ -Right)
ND5.3	5'-T CAGTTCTTCAAATAT-3' <i>De novo</i>	5'-T AAGATTAGTATGGT-3' <i>De novo</i>
ND6	5'-T GACCCCCATT-3' ^b From Bacman. S <i>et al.</i> , 2013	5'-C GATGGCTATT-3' ^c Adapted from Bacman. S <i>et al.</i> , 2013 to recognize C over T at the N_0 position
ATP8	5'-T ATTAAACACAAACTAT-3' ^d From Bacman. S <i>et al.</i> , 2013 (previously denoted as $\Delta 5$ -Left)	5'-T ATGGGCTTGGT-3' <i>De novo</i>

^{a b, c, d} The authors designed the TALE to contain a mismatched terminal RVD^{9,69}.

(See the provided Excel file).

Supplementary Table 5 | Indel frequencies of DdCBEs in their optimized orientations.

Shown are the percent of indels in HEK293T cells for each DdCBE in its optimized split orientation (see **Fig. 4a-g** for on-target editing efficiencies of optimized DdCBEs). Cells treated with *ND5.1-*, *ND5.2-*, and *ATP8-DdCBE* were harvested 3 days post-transfection; cells treated with *ND1-*, *ND2*, *ND4-* and *ND5.3-DdCBE* were harvested 6 days post-transfection.

(See the provided Excel file).

Supplementary Table 6 | Base percentages at each position of the H-strand of mtDNA for indicated DdCBE in its optimized split orientation. TALE-binding sites are highlighted in blue. On-target C•G bases are highlighted in yellow. Base percentages are measured 3 days post-transfection of HEK293T cells.

(See the provided Excel file).

Supplementary Table 7 | P-values from comparison of editing efficiencies from time course experiments in HEK293T cells and U2OS cells. For a given DdCBE, the *P*-value for editing efficiencies of target cytidine across two cumulative timepoints is shown. *P*-values were calculated using the Student's two-tailed paired t-test. Entries are highlighted in red if the *P*-value indicated a significant difference (*P*<0.05).

(See the provided Excel file).

Supplementary Table 8 | Unique off-target SNVs mediated by DdCBEs. SNVs called by VarScan 2 were considered high-confidence if the percentage frequency of a given SNV is >0.1% in one or more replicates. For samples treated with DdCBEs, dead DdCBEs, and TALE-free G1397 DddA_{tox}, the combined number of unique off-target SNVs from all three independent biological replicates that are absent in the untreated control are shown. For the untreated control, heteroplasmic mutations were excluded. SNV positions for each DdCBE treatment were from the NC_012920 reference genome. On-target SNVs are highlighted in red.

(See the provided Excel file).

Supplementary Table 9 | Overlapping off-target SNVs. Shown are the list of overlapping off-target SNVs between *ND6*- and *ND5.1*-DdCBE (yellow), *ND6*-, *ND5.1*- and *ATP8*-DdCBE (purple) and *ND6*-, *ND5.1*-, *ND5.2*- and *ATP8*-DdCBE (cyan). The average frequency of each SNV is shown.

Supplementary Table 10 | Disease-associated mitochondrial DNA point mutations. 83 pathogenic mtDNA point mutations were obtained from the MITOMAP database⁷² (accessed Dec 10, 2019). Disease-associated mutations in rRNA/tRNA and coding/non-coding regions were considered only if they had been assigned ‘Cfrm’ statuses. Blue: SNPs that are corrected by C•G-to-T•A conversions; Purple: SNPs that are corrected by A•T-to-G•C conversions; Pink: SNPs that are corrected by A•T-to-C•G conversions; Orange: SNPs that are corrected by C•G-to-G•C conversions; Yellow: SNPs that are corrected by A•T-to-T•A conversions

Allele SNP	Locus	Associated disease
T616C	MT-TF	Maternally inherited epilepsy / kidney disease
A1555G	MT-RNR1	DEAF; autism spectrum intellectual disability; possibly antiatherosclerotic
A1630G	MT-TV	MNGIE-like disease / MELAS
A3243G	MT-TL1	MELAS / LS / DMDF / MIDD / SNHL / CPEO / MM / FSGS / ASD / Cardiac+multi-organ dysfunction
T3258C	MT-TL1	MELAS / Myopathy
A3260G	MT-TL1	MMC / MELAS
T3271C	MT-TL1	MELAS / DM
A3280G	MT-TL1	Myopathy
T3291C	MT-TL1	MELAS / Myopathy / Deafness+Cognitive Impairment
A3302G	MT-TL1	MM
A4300G	MT-TI	MICM
A5690G	MT-TN	CPEO+ptosis+proximal myopathy
T5728C	MT-TN	Multiorgan failure / myopathy
A7445G	MT-TS1 precursor	SNHL
A7445G	MT-CO1	SNHL
T7510C	MT-TS1	SNHL
T7511C	MT-TS1	SNHL/Deafness
A8344G	MT-TK	MERRF; Other - LD / Depressive mood disorder / leukoencephalopathy / HiCM
T8356C	MT-TK	MERRF
T8528C	MT-ATP8/6	Infantile cardiomyopathy
T8851C	MT-ATP6	BSN / Leigh syndrome
T8993C	MT-ATP6	NARP / Leigh Disease / MILS / other
T9035C	MT-ATP6	Ataxia syndromes
A9155G	MT-ATP6	MIDD, renal insufficiency
T9176C	MT-ATP6	FBSN / Leigh Disease
T9185C	MT-ATP6	Leigh Disease / Ataxia syndromes / NARP-like disease
T10010C	MT-TG	PEM
T10158C	MT-ND3	Leigh Disease / MELAS
T10191C	MT-ND3	Leigh Disease / Leigh-like Disease / ESOC
T10663C	MT-ND4L	LHON
T12706C	MT-ND5	Leigh Disease
T13094C	MT-ND5	Ataxia+PEO / MELAS, LD, LHON, myoclonus, fatigue
A13514G	MT-ND5	Leigh Disease / MELAS / Ca2+ downregulation
T14484C	MT-ND6	LHON
T14487C	MT-ND6	Dystonia / Leigh Disease / ataxia / ptosis / epilepsy
A14495G	MT-ND6	LHON
T14674C	MT-TE	Reversible COX deficiency myopathy
T14709C	MT-TE	MM+DMDF / Encephalomyopathy / Dementia+diabetes+ophthalmoplegia
T14849C	MT-CYB	EXIT / Septo-Optic Dysplasia

T14864C	MT-CYB	MELAS
A15579G	MT-CYB	Multisystem Disorder, EXIT
G583A	MT-TF	MELAS / MM & EXIT
C1494T	MT-RNR1	DEAF
G1606A	MT-TV	AMDF
G1644A	MT-TV	LS / HCM / MELAS
C3256T	MT-TL1	MELAS; possible atherosclerosis risk
C3303T	MT-TL1	MMC
G3376A	MT-ND1	LHON MELAS overlap
G3460A	MT-ND1	LHON
G3635A	MT-ND1	LHON
G3697A	MT-ND1	MELAS / LS / LDYT / BSN
G3700A	MT-ND1	LHON
G3733A	MT-ND1	LHON
G3890A	MT-ND1	Progressive Encephalomyopathy / LS / Optic Atrophy
G4298A	MT-TI	CPEO / MS
G4308A	MT-TI	CPEO
G4332A	MT-TQ	Encephalopathy / MELAS
G4450A	MT-TM	Myopathy / MELAS
G5521A	MT-TW	Mitochondrial myopathy
G5650A	MT-TA	Myopathy
G5703A	MT-TN	CPEO / MM
G7497A	MT-TS1	MM / EXIT
G8340A	MT-TK	Myopathy / Exercise Intolerance / Eye disease+SNHL
G8363A	MT-TK	MICM+DEAF / MERRF / Autism / LS / Ataxia+Lipomas
G8969A	MT-ATP6	Mitochondrial myopathy, lactic acidosis and sideroblastic anemia (MLASA) / IgG nephropathy
G10197A	MT-ND3	Leigh Disease / Dystonia / Stroke / LDYT
G11778A	MT-ND4	LHON / Progressive Dystonia
G12147A	MT-TH	MERRF-MELAS / Encephalopathy
G12276A	MT-TL2	CPEO
G12294A	MT-TL2	CPEO / EXIT+Ophthalmoplegia
G12315A	MT-TL2	CPEO / KSS / possible carotid atherosclerosis risk, trend toward myocardial infarction risk
G12316A	MT-TL2	CPEO
G13042A	MT-ND5	Optic neuropathy/ retinopathy/ LD
G13051A	MT-ND5	LHON
G13513A	MT-ND5	Leigh Disease / MELAS / LHON-MELAS Overlap Syndrome / negative association w Carotid Atherosclerosis
G14459A	MT-ND6	LDYT / Leigh Disease / dystonia / carotid atherosclerosis risk
G14710A	MT-TE	Encephalomyopathy + Retinopathy
C4171A	MT-ND1	LHON / Leigh-like phenotype
C1177A	MT-ND4	Leigh Disease
C12258A	MT-TS2	DMDF / RP+SNHL
C14482A	MT-ND6	LHON
C14482G	MT-ND6	LHON
A3243T	MT-TL1	MM / MELAS / SNHL / CPEO

Supplementary Table 11 | List of bacterial strains and plasmids used in this study

Bacterial species and strains	Genotype	Purpose	Source
<i>Escherichia coli</i> DH5α	F- φ 80/ <i>lacZΔM15</i> Δ(<i>lacZYA-argF</i>)U169 <i>recA1 endA1 hsdR17(rK-, mK+) phoA supE44 λ-thi-1 gyrA96 relA1</i>	General cloning	Thermo Fisher Scientific Cat#18258012
<i>Escherichia coli</i> DH5α: <i>dddI</i>	F- φ 80/ <i>lacZΔM15</i> Δ(<i>lacZYA-argF</i>)U169 <i>recA1 endA1 hsdR17(rK-, mK+) phoA supE44 λ-thi-1 gyrA96 relA1, with chromosomally integrated <i>dddI</i></i>	Cloning for full-length <i>DddA_{tox}</i> fusions	This study
<i>Escherichia coli</i> Mach1	Δ <i>recA1398 endA1 fhuA</i> Φ 80Δ(<i>lac</i>)M 15 Δ(<i>lac</i>)X74 <i>hsdR(rK-mK+)</i>	General cloning (mammalian plasmids)	Thermo Fisher Scientific Cat# C862003
<i>Escherichia coli</i> BL21	F- <i>ompT hsdS_B(r_B⁻, m_B⁻) gal dcm</i> (DE3)	Protein extraction	EMD Millipore Cat#69450
<i>Escherichia coli</i> XK1502	F- Δ <i>lacU169 nalA</i>	Protein expression	73
<i>Burkholderia cenocepacia</i> H111	Wild type	Amplification of <i>dddA</i> - <i>dddA_I</i> , parental strain for mutant construction	74
<i>Burkholderia cenocepacia</i> H111 <i>ΔicmF1</i> GentR	Δ I35_RS01770, attTn7::aacC1	Growth competition assays	This study
<i>Burkholderia cenocepacia</i> H111 <i>ΔicmF2</i> GentR	Δ I35_RS17395, attTn7::aacC1	Growth competition assays	This study
<i>Burkholderia cenocepacia</i> H111 <i>ΔdddAΔdddA_I</i> GentR	ΔI35_RS34180Δ I35_RS34175, attTn7::aacC1	Growth competition assays	This study
<i>Burkholderia cenocepacia</i> H111 <i>dddAE1347A</i>	I35_RS34180 ^{E1347A} , attTn7::aacC1	Growth competition assays	This study
Plasmids	Purpose	Source	
pPSV39-CV	For inducible expression of proteins in <i>E. coli</i>	75	
pScrhaB2-V	For inducible expression of proteins in <i>E. coli</i>	76	
pBAD24	Protein expression	77	
pETDuet-1	Protein expression	EMD Millipore Cat#71146-3	
pDONRPEX18Gm-SceI- <i>pheS</i>	Allelic replacement in <i>B. cenocepacia</i>	78	
pDAI-SceI- <i>pheS</i>	Allelic replacement in <i>B. cenocepacia</i>	78	
pUC18T-mini-Tn7T-Gm	For the generation of gentamycin resistant <i>B. cenocepacia</i>	79	
pScrhaB2-V:: <i>dddA</i>	To express <i>dddA</i>	This study	
pScrhaB2-V:: <i>cdd</i>	To express <i>cdd</i>	This study	
pScrhaB2-V::APOBEC3G	To express APOBEC3G	This study	
pScrhaB2-V:: <i>tadA</i>	To express <i>tadA</i>	This study	
pPSV39-CV:: <i>dddA_I</i>	To express <i>dddA_I</i>	This study	
pDONRPEX18Gm-SceI- <i>pheS</i> :: <i>ΔicmF1</i>	To delete <i>icmF1</i> from <i>B. cenocepacia</i>	This study	
pDONRPEX18Gm-SceI- <i>pheS</i> :: <i>ΔicmF2</i>	To delete <i>icmF2</i> from <i>B. cenocepacia</i>	This study	
pDONRPEX18Gm-SceI- <i>pheS</i> :: <i>ΔdddAΔdddA_I</i>	To delete <i>dddA</i> and <i>dddA_I</i> from <i>B. cenocepacia</i>	This study	
pDONRPEX18Gm-SceI- <i>pheS</i> :: <i>dddA^{E1347A}</i>	To generate the <i>dddA^{E1347A}</i> catalytic mutant in <i>B. cenocepacia</i>	This study	
pBAD24:: <i>ung</i>	To express <i>ung</i> in <i>E. coli</i>	This study	

pETDuet-1 mcs1:: <i>dddA-his₆</i> <i>mcs1::dddA₁</i>	To co-express <i>dddA-dddA₁</i>	This study
pTNS3	Tn7 transposase expression	80
pRK2013	Helper plasmid for plasmid mobilization	ATCC® 37159
pKD46	Recombinase expression plasmid for λ Red deletion system	81
pKD4	Template for <i>ung</i> deletion amplicon	81
pCMV	For mammalian expression of BE2, BE4max, split DddA _{tox} and all DddA _{tox} fusions	14, 82

Supplementary Table 12 | Sequences used for DddA_{tox} characterization in bacteria

Primers for cloning *dddA* deaminase domain into pScrhaB2-V

NdeI-DddA TCAAGTACTACATATGATAGGACTCAACGGTGGGC
DddA-NS- TACTGATTGATCTAGAACAAACCTCCTTCGTGGGG
XbaI

Primers for cloning *dddA*, deaminase domain into pPSV39-CV

DddAI-XbaI TACTGATTGATCTAGATTACAACACTCGCTCCATGTCAGTTG
Sacl-RBS- TCAAGTACTAGAGCTCACGGGAGGAAAGATGTACGCAGACGA
DddAI TTTCGACG

Primers for cloning *dddA* deaminase domain into pETDuet-1 mcs1

BamHI- TATCAGAAACGGATCCATAGGACTCAACGGTGGGC
DddA_duet
DddA_duet- TATGTTACTAGCGGCCGCTCAACAAACCTCCTTCGTGGG
NotI

Primers for cloning *dddA*, deaminase domain into pETDuet-1 mcs1

NdeI_DddAI TCAAGTACTACATATGTCAGCAGACGATTCGACG
DddAI-BglII TATGTTACTAAGATCTTACAACACTCGCTCCATGTCAGTTG

Primers for cloning *dddA(E1347A)* deaminase domain into pETDuet-1 mcs1

DddA_E134 CGGACTGACCGGCAACGTGCCGGCGTTGC
7A-3
DddA_E134 GGCACGTTGCCGGTCAGTCCGCCTATTATGC
7A-4

Primers for construction of deletion cassette for *icmF1* using pDONRPEX18Gm-SceI-pheS

HindIII- TGTAAAGCTAAAGCTTAGGGATAACAGGGTAATCTGCTGGAT
ISceI-IcmF1 CCGGATTTCCG
IcmF1-2 TTCAGCATGCTTGCAGCTCGAGTTGATGCGTTGCATAGGACGT
TCA
Icmf1-3 AACTCGAGCCGCAAGCATGCTGAAAGGGCGCAATGACGAAA
CC
Bcen_IcmF1 TCAATCAGTATCTAGAGTAGAACGGATCGACCGGCA
-XbaI
HindIII- TGTAAAGCTAAAGCTTAGGGATAACAGGGTAATCGCTCATTG
ISceI-IcmF2 TCCGTTGCAGC
IcmF2-2 TTCAGCATGCTTGCAGCTCGAGTTGCGAACGATCATGTGTGAT
ACAC
IcmF2-3 AACTCGAGCCGCAAGCATGCTGAATTTCGAGACCCGCGATGA
CG
IcmF2-XbaI TCAATCAGTATCTAGACGAGCCGCTCGATACGATTG

Primers for construction of deletion cassette for *dddA*-*dddA*, using pDONRPEX18Gm-SceI-pheS

HindIII-1Secl-	TGTTAAGCTAAAGCTTAGGGATAACAGGGTAATGTGGTACTTCAAC
DddADddAI	GAAGCAGATG
DddADddAI-2	TTCAGCATGCTGCGGCTCGAGTTATCGGATCAGTGACTCGTGC
DddADddAI-3	AACTCGAGCCGCAAGCATGCTGAAAGCGAGTTGTAAGAACGGAGC
Bcen_E1-I1-	TCAATCAGTATCTAGAAGTGAGCTCTCCGAAATCGAAC
XbaI	

Primers for construction of *dddA*(E147A) using pDONRPEX18Gm-SceI-pheS

Bcen_E1347A-1	TAACACGACGGCCAGTGCCAAGCTAGGGATAACAGGGTAATAGCA GCTACGTGTACAGTCCGGACGCACCGTATTGC
Bcen_E1347A-2	AATAAGGCAGACTGACCGGCAACGTGCCCGGCGTTGCGTAGTTGG
Bcen_E1347A-3	CCGGGCACGTTGCCGGTCAGTCCGCCTATTATGCGC
Bcen_E1347A-4	GCTCGGTACCCGGGATCCTCTAGACTCGCTCCATGTCAGTTGCTC GGGCG

Primers for cloning of *cdd* into pScrhaB2-V

CDD_fwd	TGAAATTCAAGCAGGATCACATATGCATCCACGTTTCAAACCGC
CDD_rev	TGCATGCCTGCAGGTCGACTCTAGAAGCGAGAAGCACTCGGTC

Primers for cloning of *tadA* into pScrhaB2-V

tadA_fwd	AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGC TCTTCCGATCT
tadA_rev	TGCATGCCTGCAGGTCGACTCTAGACTAAAAATACGTATCGCTTA G

Primers for cloning of *A3G* into pScrhaB2-V

A3G_fwd	TGAAATTCAAGCAGGATCACATATGGACCCACCAACTTTACTTTAAT TTAAC
A3G_rev	TGCCTGCAGGTCGACTCTAGAGTTTCTGGTTTGAAGATTG

Primers for *ung* deletion in *E.coli*

ung_del-F	TAGAAAGAAGCAGTTAAGCTAGGCGGATTGAAGATTGCGAGGAGAG CGAGATGGTAGGCTGGAGCTGCTTC
ung_del-R	TGATAAAATCAGCCGGTGGCAACTCTGCCATCCGGCATTCCCCGC AAATTACATATGAATATCCTCCTAG

Primers for cloning *ung* into pBAD24

EcoRI_ung	TAGTACAGAGAATTGCTAACGAATTACCTGGCATGAC
ung_XbaI	TCAATCAGTATCTAGATTACTCACTCTGCCGGTAATACTG

Sequences for poisoned primer extension assay

Substrate	
GUUG	AUGGUUGGUAGUGGAUGUGGAUAAGAUGGAG
Substrate	
GUCG	AUGGUCGGUAGUGGAUGUGGAUAAGAUGGAG
Primer	5'FAM-CTCCATCTTATCCACATCCACT

Supplementary Table 13 | Sequences used for characterization of DddA_{tox} and its fusions in mammalian cells.

Primers used for generating sgRNA transfection plasmids. Rev_sgRNA_plasmid was used in all cases

rev_sgRNA_plasmid	GGTGTTCGTCCTTCCACAAG
fwd_saG1	GTCTGTGCCCTCCCTCCCTGGCGTTTAGTACTCT
fwd_saG2	GTAATGAAAATTACAGAACTAC
fwd_saG3	GCCCCTCCCTCCCTGGCCCAGGTGTTAGTACTCT
fwd_saG4	GTAATGAAAATTACAGAACTAC
fwd_spG6	GCCCTCCCTGGCCCAGGTGAAGGGTTAGTACTC
fwd_spG7	TGTAATGAAAATTACAGAACTAC
fwd_EMX1	GTGTGGTCCAGAACCGGAGGAGTTAGTACTCTG
	TAATGAAAATTACAGAACTAC
	GAGGCCCGAGAGCAGCCACGTTAGAGCTAGAA
	ATAGCAAGTAAAATAAGGC
	GCCACTGGGCCTCAACACTCGTTAGAGCTAGAA
	ATAGCAAGTAAAATAAGGC
	GAGTCCGAGCAGAAGAAGTTAGAGCTAGAAA
	TAGCAAGTAAAATAAGGC

Primers for HTS of on-target sites from all mammalian cell culture experiments

fwd_CCR5HTS	ACACTCTTCCCTACACGACGCTCTCCGATCTNNN NCAAGTGTGATCACTGGGTGG
rev_CCR5HTS	TGGAGTTCAGACGTGTGCTCTCCGATCTGGATTCC CGAGTAGCAGATG
fwd_EMX1HTS	ACACTCTTCCCTACACGACGCTCTCCGATCTNNN NGGCCCTAACCTATGTAGC
rev_EMX1HTS	TGGAGTTCAGACGTGTGCTCTCCGATCTCTCTGC TCGGACTCAGGC
fwd_ND1HTS	ACACTCTTCCCTACACGACGCTCTCCGATCTNNN NCTCACCATCGCTCTTCTACTATG
rev_ND1HTS	TGGAGTTCAGACGTGTGCTCTCCGATCTGGCTAG GGTGAATTATGAG
fwd_ND2HTS	ACACTCTTCCCTACACGACGCTCTCCGATCTNNN NCGTAAGCCTCTCCTCACT
rev_ND2HTS	TGGAGTTCAGACGTGTGCTCTCCGATCTGTTGAGT AGTAGGAATGCGGTAG
fwd_ND4HTS	ACACTCTTCCCTACACGACGCTCTCCGATCTNNN NGACTTCAAACCTACTCCCACAAATAG
rev_ND4HTS	TGGAGTTCAGACGTGTGCTCTCCGATCTGTTGAGT TAAATATGTAGAGGGAG
fwd_ND5.1/5.2HTS	ACACTCTTCCCTACACGACGCTCTCCGATCTNNN NCGGGTCCATCATCCACAAAC
rev_ND5.1/5.2HTS	TGGAGTTCAGACGTGTGCTCTCCGATCTAGAGTAA TAGATAGGGCTCAGGC

fwd_ND5.3HTS	ACACTCTTCCCTACACGACGCTCTCCGATCTNNN NCTGTAGCATTGTCGTTACATGG
rev_ND5.3HTS	TGGAGTTCAGACGTGTGCTCTCCGATCTCTGATGA GCAAGAAGGATATAATTCC
fwd_ND6HTS	ACACTCTTCCCTACACGACGCTCTCCGATCTNNN NCTCTTCACCCACAGCACC
rev_ND6HTS	TGGAGTTCAGACGTGTGCTCTCCGATCTGATTGTT AGCGGTGTGGTCG
fwd_ATP8HTS	ACACTCTTCCCTACACGACGCTCTCCGATCTNNN NCTTTACAGTCAAATGCCCAAC
rev_ATP8HTS	TGGAGTTCAGACGTGTGCTCTCCGATCTGGGGGC AATGAATGAAGCG

Primers for HTS of off-target sites from all mammalian cell culture experiments

fwd_5303HTS	ACACTCTTCCCTACACGACGCTCTCCGATCTNNN NGCTAACATGACTAACACCCCTTAATTTC
rev_5303HTS	TGGAGTTCAGACGTGTGCTCTCCGATCTGTAGGAG TAGCGTGGTAAGG
fwd_7994/8115HTS	ACACTCTTCCCTACACGACGCTCTCCGATCTNNN NCCCCCATTATTCCTAGAACAG
rev_7994/8115HTS	TGGAGTTCAGACGTGTGCTCTCCGATCTGCTATAG GGTAAATACGGGCC
fwd_8619/8648/8720HTS	ACACTCTTCCCTACACGACGCTCTCCGATCTNNN NGATCATTCTATTTCCCCCTCTATTG
rev_8619/8648/8720HTS	TGGAGTTCAGACGTGTGCTCTCCGATCTCACTGTG CCCGCTCATAAG
fwd_10192/10205/10349HTS	ACACTCTTCCCTACACGACGCTCTCCGATCTNNN NGAAAAAATCCACCCCTTACGAG
rev_10192/10205/10349HTS	TGGAGTTCAGACGTGTGCTCTCCGATCTCGTTTG TTTAAACTATATACCAATTCGG
fwd_13763HTS	ACACTCTTCCCTACACGACGCTCTCCGATCTNNN NCCCCACCCCTACTAACATTAACG
rev_13763HTS	TGGAGTTCAGACGTGTGCTCTCCGATCTGTTGTT GGTTAGGTAGTTGAGG
fwd_15598/15619/15646/15675HTS	ACACTCTTCCCTACACGACGCTCTCCGATCTNNN NCCTATTGCGCTACACAATTCTC
rev_15598/15619/15646/15675HTS	TGGAGTTCAGACGTGTGCTCTCCGATCTGTTGGTA TTAGGATTAGGATTGTTGTG
fwd_15950HTS	ACACTCTTCCCTACACGACGCTCTCCGATCTNNN NCTCAAATGGGCCTGTCCTTG
rev_15950HTS	TGGAGTTCAGACGTGTGCTCTCCGATCTGTACTAC AGGTGGTCAAGTATTATG
fwd_16363/16393/16394HTS	ACACTCTTCCCTACACGACGCTCTCCGATCTNNN NCCATTACCGTACATAGCACATTAC
rev_16363/16393/16394HTS	TGGAGTTCAGACGTGTGCTCTCCGATCTCGTGAGT GGTTAATAGGGTGATAG

Primers for mtDNA copy number analysis by quantitative PCR

fwd_ND5.1/5.2_qPCR	GAACAAGATATTGAAAAATAGGAGGAC
rev_ND5.1/5.2_qPCR	GCGGTTTCGATGATGTGG
fwd_ND6_qPCR	CACTCACCAAGACCTCAACC
rev_ND6_qPCR	GAATGATGGTTGTCTTGGATATACTAC
fwd_ATP8_qPCR	CTTACACTATTCCCTCATCACCCAAC
rev_ATP8_qPCR	GTTCATTTGGTTCTCAGGGTTG
fwd_β-actin_qPCR	AGGCACCAGGGCGTGAT
rev_β-actin_qPCR	CAGGGTGAGGATGCCTC

Primers for long-range PCR of whole mitochondrial genome as two amplicons

fwd_2478-10858	GCAAATCTTACCCCGCCTG
rev_2478-10858	AATTAGGCTGTGGGTGGTTG
fwd_2688-10653	GCCATACTAGTCTTGCCGC
rev_2688-10653	GGCAGGTCAATTCACTGG

Primer for amplification of ND4 gene fragement from ND4-edited cells

fwd_ND4_PCR	GCCATTCTCATCCAAACC
Rev_ND4_PCR	GGTGAGGGATAGGAGGAG

Primers for qPCR and RT-qPCR of ND4-edited cells

fwd_ND4_RT-qPCR	CAAGCTCCATCTGCCTACGA
rev_ND4_RT-qPCR	GCGATTATGAGAATGACTGC
fwd_RNR1_RT-qPCR	ATTACACATGCAAGCATCCC
rev_RNR1_RT-qPCR	CACGAAATTGACCAACCTG
fwd_ND1_RT-qPCR	TAGCAGAGACCAACCGAAC
rev_ND1_RT-qPCR	ATGAAGAATAGGGCGAAGGG
fwd_ND2_RT-qPCR	CTATCTCGCACCTGAAACAAGC
rev_ND2_RT-qPCR	GGTGGAGTAGATTAGGCGTAGG
fwd_ATP6/8_RT-qPCR	TGTTAGCGGTTAGGCGTA
rev_ATP6/8_RT-qPCR	TTACACCAACCACCCAAC
fwd_CO3_RT-qPCR	TTTACCCCTCCTACAAGCC
rev_CO3_RT-qPCR	GCGGATGAAGCAGATAGT
fwd_CYB_RT-qPCR	GCCTGCCTGATCCTCCAAAT
rev_CYB_RT-qPCR	AAGGTAGCGGATGATTCAAGCC
fwd_B2M_RT-qPCR_qPCR	CAGGTACTCCAAAGATTCAAGG
rev_B2M_RT-qPCR_qPCR	GTCAACTCAATGTCGGATGG

Nextera primers for ATAC-seq

i5_common	AATGATACGGCGACCACCGAGATCTACACTCGTCG
i7_1	GCAGCGTCAGATGTG
i7_2	CAAGCAGAAGACGGCATACGAGAT
	TCGCCTTAGTCTCG TGGGCTCGGAGATGT

i7_3	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTC TCGTGGGCTCGGAGATGT
i7_4	CAAGCAGAAGACGGCATACGAGATTCTGCCTGTCT CGTGGGCTCGGAGATGT
i7_5	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTC TCGTGGGCTCGGAGATGT
i7_6	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTC TCGTGGGCTCGGAGATGT
i7_7	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTC TCGTGGGCTCGGAGATGT
i7_8	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTC TCGTGGGCTCGGAGATGT
i7_9	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTC TCGTGGGCTCGGAGATGT

Supplementary sequences 1 | Intact DddA_{tox}–Cas9 fusions sequences

All intact DddA_{tox}–Cas9 have the general architecture of (from N- to C-terminus): bpNLS–DddA_{tox}–linker 1–dSpCas9 or SpCas9(D10A)–10aa linker—10aa linker– UGI–4aa linker–bpNLS

DddA_{tox}

GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGGPTYPNYANAGHVEGQSALFMRDN
GISEGLVFHNNEGTGFCVNMTETLLPENAKMTVVPPGEAIKVKGATGETKVFTGNSNSPKSPTKG
GC

dSpCas9

DKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEAKVDDSSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH
LRKKLVDSTDKAIDLRLIYLALAHMIKFRGHFLIEGDLNPNSDVKLFIQLVQTYNQLFEENPINASG
VDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDD
DLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ
QLPEKYKEIFFDQSCKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVVKLNREDILLRKQRTFDNGS
IPHQIHLGELHAILRRQEDFYPFLKDNRKIEKILTFRIPYYVGPLARGNSRFAMTRKSEETITPWN
FEEVVDKGASAQSFIERMTNFDKLPNEKVLPKHSLLYEYFTVYNELTKVKVYVTEGMRKPAFLSGEQK
KAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENE
DILEDIVLTTLFEDREMIEERLKTYAHLFDDKVMQQLKRRRTGWGRSLRKLINGIRDQSGKTILD
FLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAICKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQIQLNEKLYLYYLQ
NGRDMDVQELDINRLSDYDVDAIVPQSQLKDDSIDNKVLTRSDKNGKSDNVPSEEVVKKMKNYWRQ
LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKQLVETRQITKVAQIILDSRMNTKYDENDKLIREV
KVITLKSCLVSDFRKDFQFYKREINNYHHAHDAYLNAVVTALIKKPYLESEFVYGDYKVDVRKM
IAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWPKKYGGFDSPTVAYSVLVVAKEGKSKKL
KSVKELLGITMERSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLENGRKMLASAGELOKGNE
LALPSKYVNFLYFLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSA
YNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRID
LSQLGGD

SpCas9(D10A)

DKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEAKVDDSSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH
LRKKLVDSTDKAIDLRLIYLALAHMIKFRGHFLIEGDLNPNSDVKLFIQLVQTYNQLFEENPINASG
VDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDD
DLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ
QLPEKYKEIFFDQSCKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVVKLNREDILLRKQRTFDNGS
IPHQIHLGELHAILRRQEDFYPFLKDNRKIEKILTFRIPYYVGPLARGNSRFAMTRKSEETITPWN
FEEVVDKGASAQSFIERMTNFDKLPNEKVLPKHSLLYEYFTVYNELTKVKVYVTEGMRKPAFLSGEQK
KAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENE
DILEDIVLTTLFEDREMIEERLKTYAHLFDDKVMQQLKRRRTGWGRSLRKLINGIRDQSGKTILD
FLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAICKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQIQLNEKLYLYYLQ
NGRDMDVQELDINRLSDYDVDHIVPQSQLKDDSIDNKVLTRSDKNGKSDNVPSEEVVKKMKNYWRQ
LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKQLVETRQITKVAQIILDSRMNTKYDENDKLIREV
KVITLKSCLVSDFRKDFQFYKREINNYHHAHDAYLNAVVTALIKKPYLESEFVYGDYKVDVRKM

IAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWPKKYGGFDSPTVAYSVLVAKVEKGKSKKL
KSVKELLGITIMERSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMILASAGELQKGNE
LALPSKYVNFLYLA SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIEQISEFSKRVILADANLDKVLSA
YNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ SITGLYETRID
LSQLGGD

UGI

TNLSDIIEKETGKQLVIQESILMLPEEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWA
LVIQDSNGENKIKML

bpNLS

KRTADGSEFEPKKKRKV

Linker 1: 32aa linker

SGGSSGGSSGSETPGTSESATPESSGGSSGGS

Linker 1: 10aa flexible

GGGGSGGGGS

Linker 1: 10aa rigid

EAAAKEAAAK

Linker 1: 5aa rigid

EAAAK

10aa linker

SGGSSGGSGGS

4aa linker

SGGS

Supplementary sequences 2 | Split-DddA_{tox}–Cas9 fusions sequences

All split-DddA_{tox}–Cas9 have the general architecture of (from N- to C-terminus): **bpNLS**–DddA_{tox} **half**–32aa linker–**dSpCas9 or SaKKH-Cas9(D10A)**–10aa linker–**UGI**–10aa linker–**UGI**–4aa linker–**bpNLS**

G1333 DddA_{tox}-N

GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGG

G1333 DddA_{tox}-C

PTPYPNYANAGHVEGQSALFMRDNGI**SEGLVFHNNEGTGFCVNMTETLLPENAKMTVVPPEGAI**PV
KRGATGETKVFTGNSNSPKSPTKGCC

G1397 DddA_{tox}-N

GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGGPTPYPNYANAGHVEGQSALFMRDN
G**SEGLVFHNNEGTGFCVNMTETLLPENAKMTVVPPEG**

G1397 DddA_{tox}-C

AIPVKRGATGETKVFTGNSNSPKSPTKGCC

dSpCas9

DKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEAKVDDSFHRLIESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH
LRKKLVDSTDKAIDLRIYLALAHMIKFRGHFLIEGDLNPNSDVKLFIQLVQTYNQLFEENPINASG
VDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKNFDLAEDAKLQLSKDTYDD
DLDNLLAQIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ
QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIFKPILEKMDGTEELLVKLNRDILLRKQRTFDNGS
IPHQIHLGELHAILRRQEDFYPLKDNRREKIEKILTFRIPYYVGPLARGNSRFAMTRKSEETITPWN
FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVVKVTEGMRKPAFLSGEQK
KAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKI IKDKDFLDNEENE
DILEDIVLTTLFEDREMIEERLKYAHLFDDKVMQQLKRRRTGWGRRLSRKLINGIRDQSGKTILD
FLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQ
NGRDMDYVDQELDINRLSDYDVDA**IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMNYWRQ**
LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKQLVETRQITKHVAQILDLSRMNTKYDENDKLIREV
KVITLKSCLVSDFRKDFQFYKVREINNYHAAHDAYLNAVVTALIKKPYLESEFVYGDYKVDVRKM
IAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWPKKYGGFDSPTVAYSVLVAKVEKGKSKKL
KSVKELLGITMERSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFYFLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIEQISEFSKRVILADANLDKVLSA
YNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSIITGLYETRID
LSQLGGD

SaKKH-Cas9(D10A)

GKRNYILGLAIGITSGVGIGIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVK
KLLFDYNLLTDHSELGGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNNEVEEDTGNELSTKE
QISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSVDYVKEAKQLLKVKQKAYHQLDQSFIDTYIDL
LETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDEN
EKLEYYEKFQIIENVFKQKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVKYHDIKDTARKE
IIENAEELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELW

HTNDNQIAIFNRLKLVPKKVDSLQQKEIPTTLVDDFILSPVVKRSFIQSIVINAIKKYGLPNDIII
ELAREKNSKDAQKMINEMQKRNQQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLE
DLLNNPFNYEVDHIIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLA
KGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNDVKVKSINGGF
TSFLRRKWFKKERNKGYKHAAEDALITIANADFIFKEWKLDKAKKVMENQMFEEKQAESMPEIETEQ
EYKEIFITPHQIKHIKDFKDYKSHRVDDKPNRKLINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKL
KKLINKSPEKLLMYHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYG
NKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKK
LKKISNQAEFIASFYKNDLIKINGELYRVIGVNNNDLLNRIEVNMIDITYREYLENMNDKRPHIICKTI
ASKTQSIKKYSTDILGNLYEVSKKHPQIICKGGSPKKRKVSSDYKDHDGDYKDHDIDYKDDDDK

UGI

TNLSDIIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWA
LVIQDSNGENKIKML

bpNLS

KRTADGSEFEPKKKRKV

32aa linker

SGGSSGGSSGSETPGTSESATPESSGGSSGGS

10aa linker

SGGSGGSGGS

4aa linker

SGGS

Supplementary sequences 3 | Split-DddA_{tox}–CCR5 TALE fusions sequences

All split-DddA_{tox}–CCR5 TALE fusions have the general architecture of (from N- to C-terminus): **NLS**–**CCR5 TALE**–2aa linker– **DddA_{tox} half**–4aa linker–**UGI**

TALE domains⁸³ are annotated as: bold for N-terminal domain, underlined for RVD and bolded italics for C-terminal domain.

2aa linker
GS

CCR5 Left TALE

VDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA
**THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTQLLKIAKRGGVTAVEAVHAWRNALTGAPLN
LTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPEQVVVAIASHDGGKQALETVQRLLPVLCQAHG
LTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVVAIASNGGGKQALETVQRLLPVLCQDHG
LTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPEQVVVAIASNIGGKQALETVQRLLPVLCQAHG
LTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVVAIASNIGGKQALETVQRLLPVLCQDHG
LTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVVAIASHDGGKQALETVQRLLPVLCQAHG
LTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVVAIASHDGGKQALETVQRLLPVLCQDHG
LTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVVAIASNIGGKQALETVQRLLPVLCQAHG
LTPDQVVAIASANNGGKQALETVQRLLPVLCQAHGLTPAQVVVAIASHDGGKQALETVQRLLPVLCQDHG
LTPEQVVAIASNGGRPALE**SIVAQLSRPDPA**LAALTNDHLVALAC**LGG**RPALDAVKKGLPHAPALIK
RTNRRIPERTSHRVA******

CCR5 Right TALE

VDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA
**THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTQLLKIAKRGGVTAVEAVHAWRNALTGAPLN
LTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVVAIASNGGGKQALETVQRLLPVLCQAHG
LTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVVAIASHDGGKQALETVQRLLPVLCQDHG
LTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVVAIASNIGGKQALETVQRLLPVLCQAHG
LTPDQVVAIASANNGGKQALETVQRLLPVLCQAHGLTPAQVVVAIASNGGGKQALETVQRLLPVLCQDHG
LTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVVAIASANNGGKQALETVQRLLPVLCQAHG
LTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVVAIASNGGGKQALETVQRLLPVLCQAHG
LTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVVAIASHDGGKQALETVQRLLPVLCQDHG
LTPEQVVAIASANNGGRPALE**SIVAQLSRPDPA**LAALTNDHLVALAC**LGG**RPALDAVKKGLPHAPALIK
RTNRRIPERTSHRVA******

Supplementary sequences 4 | General DdCBE architecture and mitoTALE amino acid sequences

All right-side halves of DdCBEs have the general architecture of (from N- to C-terminus):
COX8A MTS–3xFLAG–mitoTALE–2aa linker–DddA_{tox} half–4aa linker–1x-UGI–ATP5B 3'UTR

All left-side halves of DdCBEs have the general architecture of (from N- to C-terminus):
SOD2 MTS–3xHA–mitoTALE–2aa linker–DddA_{tox} half–4aa linker–1x-UGI–SOD2 3'UTR

mitoTALE domains are annotated as: bold for N-terminal domain, underlined for RVD and bolded italics for C-terminal domain.

SOD2 MTS
LSRAVCCTSRLAPVLGYLGSRQKHSYPD

COX8A MTS
SVLTPLLLRGLTGSARRLPVPRAKIHSL

SOD2 3'UTR
ACCACGATCGTTATGCTGATCATACCCATAATGATCCCAGCAAGATAATGTCTGTCTTAAGATGTG
CATCAAGCCTGGTACATACTGAAAACCTATAAGGCCTGGATAATTTTGATTATTCAATTGAA
GAAACATTATTTCCAATTGTGAAGTTTGACTGTTAATAAAAGAACATCTGTCAACCACATCAAAAA
AAAAAAAAAA

ATP5B 3'UTR
ACCACGATCGTTATGCTGATCATACCCATAATGATCCCAGCAAGATAATGTCTGTCTTAAGATGTG
CATCAAGCCTGGTACATACTGAAAACCTATAAGGCCTGGATAATTTTGATTATTCAATTGAA
GAAACATTATTTCCAATTGTGAAGTTTGACTGTTAATAAAAGAACATCTGTCAACCACATCAAAAA
AAAAAAAAAA

ND6-DdCBE: Left mitoTALE–G1397-DddA_{tox}-N–1x-UGI (Note: Terminal **NG** RVD recognizes a mismatched T instead of a G in the reference genome)
MALSRAVCCTSRLAPVLGYLGSRQKHSYPDYDYAGYPYDVPDYAGYPYDVPDYAMDIADLRT
LGYSQQQQEKIKPKVRSTVAQHHEALVHGFTAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAI
VGVGKQWSGARALEALLTVAGELRGPPQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPOQ
VVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPEQ
VVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQ
VVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPEQ
VVAIASNNGGKQALETVQRLLPVLCQAHGLTPQVVAIASNGGGRPALE**SIVAQLSRPDPALAALTND**
HLVALACLGGRPALDAVKKGLGSGSYSALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKFSSGG
PTPYPNYANAGHVEGQSALFMRDNGISEGKVHNPEGTCGFCVNMTETLLPENAKMTVVPEGSGGS
TNLDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWA
LVIQDSNGENKIKML**

ND6-DdCBE: Right mitoTALE–G1397-DddA_{tox}-N–1x-UGI (Note: Terminal **NG** RVD recognizes a mismatched T instead of a G in the reference genome. The NTD was also engineered to be permissive for A, T, C and G nucleotides at the *N*₀ position)

MASVLTPLLRGLTGSARRLPVPRAKIHSLDYKDHDGDYKDHIDYKDDDKMDIADLRTLGYSQQQQ
EKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTAVKYQDMIAALPEATHEAIVGVGKRGAGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPNLTPQQVVAIASNN
 GGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPQQVVAIASNG
 GGKQALETVQRLLPVLCQAHGLTPQQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPQQVVAIASNN
 GGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPQQVVAIASNG
 GGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPQQVVAIASNG
 GGKQALETVQRLLPVLCQAHGLTPQQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPQQVVAIASNG
GRPALESIVAQSLSRPDPALAALTNDHLVALACLGGRPALDAVKKGGSAPIVKRGATGETKVFTGN
 SNSPKSPTKGCGSGSNTNLSDIIEKETGKQLVIQESILMPEEVEEVIGNKPESDILVHTAYDESTDE
 NVMLLTSDAPEYKPWALVIQDSNGENKIKML**

ND1-DdCBE Right mitoTALE repeat

DIADLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTAVKYQDMIAALP
EATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAP
LNLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPQQVVAIASNIGGKQALETVQRLLPVLCQA
 HGLTPQQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQA
 HGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQA
 HGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
 HGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQA
 HGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQA
 HGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
 HGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQA
 HGLTPQQVVAIASNNGGKQALETVQALLPVLCQAHGLTPQQVVAIASNNGGKQALETVQALLPVLCQA
HGLTPQQVVAIASNNGGKQALETVQALLPVLCQAHGLTPQQVVAIASNNGGKQALETVQALLPVLCQA
SIVAQSLSRPDPA
LAALTNDHLVALACLGGRPALDAVKKG

ND1-DdCBE Left mitoTALE repeat

DIADLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTAVKYQDMIAALP
EATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAP
LNLTPEQVVAIASHDGGKQALETVQALLPVLCQAHGLTPQQVVAIASNNGGKQALETVQRLLPVLCQA
 HGLTPQQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQA
 HGLTPEQVVAIASHDGGKQALETVQALLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQA
 HGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
 HGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQA
 HGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQA
 HGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQA
HGLTPQQVVAIASNNGGKQALETVQALLPVLCQAHGLTPQQVVAIASNNGGKQALETVQALLPVLCQA
SIVAQSLSRPDPALAALTNDHLVALACLGGRPALDAVKKG

ND2-DdCBE Right mitoTALE repeat

DIADLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTAVKYQDMIAALP
EATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAP
LNLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPQQVVAIASNNGGKQALETVQRLLPVLCQA
 HGLTPQQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
 HGLTPEQVVAIASHDGGKQALETVQALLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQA
 HGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQA
 HGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
 HGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQA
 HGLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPQQVVAIASNNGGKQALETVQALLPVLCQA
HGLTPQQVVAIASNNGGKQALETVQALLPVLCQAHGLTPQQVVAIASNNGGKQALETVQALLPVLCQA
SIVAQSLSRPDPALAALTNDHLVALACLGGRPALDAVKKG

ND2-DdCBE Left mitoTALE repeat

**DIADLRTLGYSQQQQEKKPKVVRSTVAQHHEALVGHGFTAHIVALSQHPAALGTAVKYQDMIAALP
EATHEAIVGVGKQWSGARALEALLTVAGELRGPPQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAP
LNLTPEQVVAIASHDGGKQALETVQALLPVLCQAHGLTPQQVVAIASNNGGKQALETVQRLLPVLCQA
HGLTPQQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
**SIVAQLSRPDPA
LAALTNDHLVALACLGGRPALDAVKKGLG****

ND4-DdCBE Right mitoTALE repeat

**DIADLRTLGYSQQQQEKKPKVVRSTVAQHHEALVGHGFTAHIVALSQHPAALGTAVKYQDMIAALP
EATHEAIVGVGKQWSGARALEALLTVAGELRGPPQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAP
LNLTPEQVVAIASHDGGKQALETVQALLPVLCQAHGLTPQQVVAIASNIGGKQALETVQRLLPVLCQA
HGLTPQQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQA
**SIVAQLSRPDPA
LAALTNDHLVALACLGGRPALDAVKKGLG****

ND4-DdCBE Left mitoTALE repeat

**DIADLRTLGYSQQQQEKKPKVVRSTVAQHHEALVGHGFTAHIVALSQHPAALGTAVKYQDMIAALP
EATHEAIVGVGKQWSGARALEALLTVAGELRGPPQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAP
LNLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPQQVVAIASNIGGKQALETVQRLLPVLCQA
HGLTPQQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
**SIVAQLSRPDPA
LAALTNDHLVALACLGGRPALDAVKKGLG****

ND5.1-DdCBE Right mitoTALE repeat

**DIADLRTLGYSQQQQEKKPKVVRSTVAQHHEALVGHGFTAHIVALSQHPAALGTAVKYQDMIAALP
EATHEAIVGVGKQWSGARALEALLTVAGELRGPPQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAP
LNLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNNGGKQALETVQRLLPVLCQA
HGLTPQQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQA
HGLTPQQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPQQVVAIASNNGGKQALETVQRLLPVLCQA
HGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNNGGKQALETVQRLLPVLCQA
HGLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQA
HGLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQA
HGLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQA
**SIVAQLSRPDPA
LAALTNDHLVALACLGGRPALDAVKKGLG****

ND5.1-DdCBE Left mitoTALE repeat

**DIADLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTAHIVALSQHPAALGTAVKYQDMIAALP
EATHEAIVGVGKQWSGARALEALLTVAGELRGPPQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAP
LNLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPQQVVAIASNNGGKQALETVQRLLPVLCQA
HGLTPQQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPQQVVAIASNNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNNGGGKQALETVQALLPVLCQA
HGLTPQQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNNGGGKQALETVQALLPVLCQA
HGLTPQQVVAIASNNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPQQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPQQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPQQVVAIASNNGGRPALES**SIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKG**LG**

ND5.2-DdCBE Right mitoTALE repeat (Note: Terminal **NG** RVD recognizes a mismatched T instead of a G in the reference genome)

**DIADLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTAHIVALSQHPAALGTAVKYQDMIAALP
EATHEAIVGVGKQWSGARALEALLTVAGELRGPPQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAP
LNLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQA
HGLTPQQVVAIASNNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPQQVVAIASNNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNNGGGKQALETVQALLPVLCQA
HGLTPQQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNNGGGKQALETVQALLPVLCQA
HGLTPQQVVAIASNNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQA
HGLTPQQVVAIASNNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNNGGGKQALETVQRLLPVLCQA
HGLTPQQVVAIASNNGGRPALES**SIVAQLSRPDPA**
LAALTNDHLVALACLGGRPALDAVKKGLG**

ND5.2-DdCBE Left mitoTALE repeat

**DIADLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTAHIVALSQHPAALGTAVKYQDMIAALP
EATHEAIVGVGKQWSGARALEALLTVAGELRGPPQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAP
LNLTPEQVVAIASHDGGKQALETVQALLPVLCQAHGLTPQQVVAIASNIGGKQALETVQRLLPVLCQA
HGLTPQQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPQQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPQQVVAIASNNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQA
HGLTPQQVVAIASNNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQA
HGLTPQQVVAIASNNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQA
HGLTPQQVVAIASNNGGRPALES**SIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKG**LG**

ND5.3-DdCBE Right mitoTALE repeat

**DIADLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTAHIVALSQHPAALGTAVKYQDMIAALP
EATHEAIVGVGKQWSGARALEALLTVAGELRGPPQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAP
LNLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQA
HGLTPQQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPQQVVAIASNNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNNGGGKQALETVQRLLPVLCQA
HGLTPQQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNNGGGKQALETVQRLLPVLCQA
HGLTPQQVVAIASNNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQA
HGLTPQQVVAIASNNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNNGGGKQALETVQRLLPVLCQA
HGLTPQQVVAIASNNGGRPALES**SIVAQLSRPDPA**
LAALTNDHLVALACLGGRPALDAVKKGLG**

ND5.3-DdCBE Left mitoTALE repeat

**DIADLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTAHIVALSQHPAALGTAVKYQDMIAALP
EATHEAIVGVGKQWSGARALEALLTVAGELRGPPQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAP
LNLTPEQVVAIASHDGGKQALETVQALLPVLCQAHGLTPQQVVAIASNIGGKQALETVQRLLPVLCQA
HGLTPQQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNGGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPQQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPQQVVAIASNGGGRPALES**SIVAQLSRPDPA**LAALTNDHLVALACLGGRPALDAVKKG**LG****

ATP8-DdCBE Right mitoTALE repeat (Note: Terminal **NG** RVD recognizes a mismatched T instead of a C in the reference genome)

**DIADLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTAHIVALSQHPAALGTAVKYQDMIAALP
EATHEAIVGVGKQWSGARALEALLTVAGELRGPPQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAP
LNLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQA
HGLTPQQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNGGGKQALETVQALLPVLCQA
HGLTPQQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQA
HGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPQQVVAIASNIGGKQALETVQRLLPVLCQA
HGLTPQQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNNGGGKQALETVQRLLPVLCQA
HGLTPEQVVAIASNNGGGKQALETVQRLLPVLCQAHGLTPQQVVAIAS**NGGGRPALES****SIVAQLSRPDPA**
LAALTNDHLVALACLGGRPALDAVKKGLG******

ATP8-DdCBE Left mitoTALE repeat

**DIADLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTAHIVALSQHPAALGTAVKYQDMIAALP
EATHEAIVGVGKQWSGARALEALLTVAGELRGPPQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAP
LNLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPQQVVAIASNGGGKQALETVQRLLPVLCQA
HGLTPQQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPQQVVAIASNIGGKQALETVQALLPVLCQA
HGLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPQQVVAIAS**NGGGRPALES****SIVAQLSRPDPA**
LAALTNDHLVALACLGGRPALDAVKKGLG******

Supplementary References

- 58 Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).
- 59 Cong, L. et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819-823, doi:10.1126/science.1231143 (2013).
- 60 Bothmer, A. et al. Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus. *Nature Communications* **8**, 13905, doi:10.1038/ncomms13905 (2017).
- 61 Harrison, L., Brame, K. L., Geltz, L. E. & Landry, A. M. Closely opposed apurinic/apyrimidinic sites are converted to double strand breaks in Escherichia coli even in the absence of exonuclease III, endonuclease IV, nucleotide excision repair and AP lyase cleavage. *DNA Repair (Amst)* **5**, 324-335, doi:10.1016/j.dnarep.2005.10.009 (2006).
- 62 Nissanka, N., Minczuk, M. & Moraes, C. T. Mechanisms of Mitochondrial DNA Deletion Formation. *Trends Genet* **35**, 235-244, doi:10.1016/j.tig.2019.01.001 (2019).
- 63 Andreazza, S. et al. Mitochondrially-targeted APOBEC1 is a potent mtDNA mutator affecting mitochondrial function and organismal fitness in Drosophila. *Nat Commun* **10**, 3280, doi:10.1038/s41467-019-10857-y (2019).
- 64 Lin, Y. et al. SAPTA: a new design tool for improving TALE nuclease activity. *Nucleic Acids Res* **42**, e47, doi:10.1093/nar/gkt1363 (2014).
- 65 Moscou, M. J. & Bogdanove, A. J. A simple cipher governs DNA recognition by TAL effectors. *Science* **326**, 1501, doi:10.1126/science.1178817 (2009).
- 66 Boch, J. et al. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509-1512, doi:10.1126/science.1178811 (2009).
- 67 Cong, L., Zhou, R., Kuo, Y.-c., Cunniff, M. & Zhang, F. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nature Communications* **3**, 968, doi:10.1038/ncomms1962 (2012).
- 68 Miller, J. C. et al. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* **29**, 143-148, doi:10.1038/nbt.1755 (2011).
- 69 Hashimoto, M. et al. MitoTALEN: A General Approach to Reduce Mutant mtDNA Loads and Restore Oxidative Phosphorylation Function in Mitochondrial Diseases. *Mol Ther* **23**, 1592-1599, doi:10.1038/mt.2015.126 (2015).
- 70 Ruiz-Pesini, E. et al. An enhanced MITOMAP with a global mtDNA mutational phylogeny. *Nucleic Acids Res* **35**, D823-828, doi:10.1093/nar/gkl927 (2007).
- 71 Lamb, B. M., Mercer, A. C. & Barbas, C. F., III. Directed evolution of the TALE N-terminal domain for recognition of all 5' bases. *Nucleic Acids Research* **41**, 9779-9785, doi:10.1093/nar/gkt754 (2013).
- 72 Lott, M. T. et al. mtDNA Variation and Analysis Using Mitomap and Mitomaster. *Curr Protoc Bioinformatics* **44**, 1 23 21-26, doi:10.1002/0471250953.bi0123s44 (2013).
- 73 Haft, R. J. et al. General mutagenesis of F plasmid Tral reveals its role in conjugative regulation. *J Bacteriol* **188**, 6346-6353, doi:10.1128/jb.00462-06 (2006).
- 74 Carlier, A. et al. Genome Sequence of Burkholderia cenocepacia H111, a Cystic Fibrosis Airway Isolate. *Genome Announc* **2**, doi:10.1128/genomeA.00298-14 (2014).
- 75 Silverman, J. M. et al. Haemolysin coregulated protein is an exported receptor and chaperone of type VI secretion substrates. *Mol Cell* **51**, 584-593, doi:10.1016/j.molcel.2013.07.025 (2013).
- 76 Cardona, S. T. & Valvano, M. A. An expression vector containing a rhamnose-inducible promoter provides tightly regulated gene expression in Burkholderia cenocepacia. *Plasmid* **54**, 219-228, doi:10.1016/j.plasmid.2005.03.004 (2005).
- 77 Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **177**, 4121-4130, doi:10.1128/jb.177.14.4121-4130.1995 (1995).
- 78 Fazli, M., Harrison, J. J., Gambino, M., Givskov, M. & Tolker-Nielsen, T. In-Frame and Unmarked Gene Deletions in Burkholderia cenocepacia via an Allelic Exchange System Compatible with Gateway Technology. *Appl Environ Microbiol* **81**, 3623-3630, doi:10.1128/AEM.03909-14 (2015).
- 79 Choi, K. H., DeShazer, D. & Schweizer, H. P. mini-Tn7 insertion in bacteria with multiple glmS-linked attTn7 sites: example Burkholderia mallei ATCC 23344. *Nat Protoc* **1**, 162-169, doi:10.1038/nprot.2006.25 (2006).

- 80 Choi, K. H. *et al.* Genetic tools for select-agent-compliant manipulation of *Burkholderia pseudomallei*.
Appl Environ Microbiol **74**, 1064-1075, doi:10.1128/aem.02430-07 (2008).
- 81 Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12
using PCR products. *Proc Natl Acad Sci U S A* **97**, 6640-6645, doi:10.1073/pnas.120163297 (2000).
- 82 Guilinger, J. P. *et al.* Broad specificity profiling of TALENs results in engineered nucleases with
improved DNA-cleavage specificity. *Nat Methods* **11**, 429-435, doi:10.1038/nmeth.2845 (2014).
- 83 Reyon, D. *et al.* Engineering customized TALE nucleases (TALENs) and TALE transcription factors by
fast ligation-based automatable solid-phase high-throughput (FLASH) assembly. *Curr Protoc Mol Biol*
Chapter 12, Unit 12 16, doi:10.1002/0471142727.mb1216s103 (2013).